基于深层注意力的 LSTM 的特定主题情感分析

胡朝举，梁　宁

(华北电力大学 控制与计算机工程学院, 河北 保定 071000)

摘　要：目前特定主题情感分析任务中，传统的基于注意力的深度学习模型缺乏对主题特征和情感信息的有效关注。针对该问题，构建了融合主题特征的深层注意力的 LSTM 模型（deeper attention LSTM with aspect embedding，AE-DATT-LSTM），通过共享权重的双向 LSTM 将主题词向量和文本词向量进行训练，得到主题特征和文本特征进行特征融合，经过深层注意力机制的处理，由分类器得到相应主题的情感分类结果。在 SemEval-2014 Task4 和 SemEval-2017 Task4 数据集上的实验结果表明，该方法在特定主题情感分析任务中，较之前基于注意力的情感分析模型在准确率和稳定性上有了进一步的提高。主题特征和深层注意力机制的引入，对于基于特定主题的情感分类任务具有重要的意义，为舆情分析、问答系统和文本推理等领域提供了方法的支持。

关键词：特定主题情感分析；深层注意力；LSTM；深度学习；自然语言处理

中图分类号：TP391　　doi: 10.3969/j.issn.1001-3695.2017.11.0736

Deeper attention-based LSTM for aspect sentiment analysis

Hu Chaoju, Liang Ning

(*School of Control & Computer Engineering*, *North China Electric Power University*, *Baoding Hebei* 071000, *China*)

Abstract: In the current aspect sentiment analysis task, the traditional attention-based deep learning model lacks the effective attention to the aspect information and the sentiment information. This paper put forward a new LSTM model, which combining aspect information and deeper attention. Through the bidirectional LSTM with shared weights, it trained the aspect embedding and the text embedding to get the aspect feature and text feature to carry on the feature fusion, and after the deeper attention mechanism processing, it obtained the classification result of the corresponding aspect by the classifier. The experimental results of the SemEval-2014 Task4 and SemEval-2017 Task4 datasets show that this method has further improved the accuracy and stability of the attention-based sentiment analysis model in the aspect sentiment analysis. The introduction of aspect features and deeper attention mechanisms is of great significance to the task of sentiment analysis based on aspect, which provides method support for public opinion analysis, question answering system and text reasoning.

Key words: aspect sentiment analysis; deeper attention; LSTM; deep learning; natural language processing

0　引言

随着互联网行业的发展，电商评论和社交平台的推广，积累了大量的短文本信息数据。通过提取这些数据中蕴涵的观点和情感，能够帮助人们更好地作一些决策和推广。利用自然语言处理的技术进行文本情感分析，已成为国内外研究的热点领域[1]。

情感分析，也称为观点挖掘，作为 NLP 领域和计算语言学的基本任务，近几年在工业界和学术界受到越来越多的关注。情感分析任务的关键是去理解用户在社交媒体和产品评论中产生的文本信息，从中挖掘情感信息[2]。

近年来，基于深度学习的方法在许多自然语言处理的任务中有着优异的表现，如机器翻译、语义识别、问答系统和文本摘要等。同时，在情感分析任务中，深度学习模型在没有人工标注的特征工程前提下，提高了分类的准确率。Socher 等人[3]将递归神经网络用于情感树的构建，提高了情感检测的准确率。Tang 等人[4]采用循环神经网络建立篇章级循环神经网络模型，该模型相比标准的循环神经网络模型具有较高的优越性，在情感分类任务中取得了进步。Tai 等人[5]改进标准的 LSTM 模型，引入 Tree-LSTM 模型，该模型建立了树状 LSTM 的网络拓扑结构，在情感分类任务中有较好的表现。

注意力机制来源于人脑关注事物的关键部分分配更多的注意力，通过计算注意力概率分布，对关键性输入进行突出，从而对传统的模型起到优化作用。注意力机制最早由 Mnih 等人

收稿日期：2017-11-05；**修回日期：**2018-01-12

作者简介：胡朝举（1963-），男，副教授，硕导，主要研究方向为数据库、信息安全；梁宁（1992-），男，山东泰安人，硕士研究生，主要研究方向为数据库、自然语言处理、情感分析（liangshanbo27@163.com）.

[6]将循环神经网络模型中引入注意力机制，从而提高了图像分类的准确率。随后，注意力机制在机器翻译[7]和文本摘要[8]等自然语言处理领域中取得了很好的效果。Yang 等人[9]提出的层级注意力网络用于文本分类，分类的效果要比之前提出的模型效果好。Pavlopoulos 等人[10]提出深层注意力机制，将其应用到审核用户评论，取得了比 RNN 更好的效果。在自然语言处理领域，注意力机制能够使模型关注句子中重要的部分。在情感分析任务中，许多模型已经利用注意力能够获取句子中重要信息这一优势，从而获取到比较理想的效果。

1　相关工作

1.1　特定主题情感分析

特定主题情感分类是细粒度情感分类任务，该任务的目标是对于给定的句子和句子中出现的主题，推测相应主题对应的情感极性[11, 12]，如 Positive、Negative 和 Neutral。例如，在“the food is great and they have a good selection of wines at reasonable prices.”中，主题“food”的情感极性是 Positive，然而主题“price”的情感极性是 Negative。本文中的主题信息来源于已提取好的训练集和测试集，在实际应用中，可以通过 TF-IDF 或者 LDA 来在关键词中自动提取主题信息。目前，大多数文献中的方法尝试去挖掘整个句子的情感极性，没有考虑针对特定主题的情况。传统的方法经过一系列的人工标注特征来解决这个问题。随着情感词典的不断构建，情感分析任务使用基于词典的特征进行分析。大多数的研究集中在利用特征构建情感分类器，包括词袋模型和情感词典的构建，具有代表性的方法包括基于 SVM 的方法[13]和基于神经网络的方法[14]。神经网络模型能够在不需要关注特征工程能够很好地学习到数据中的特征表示，并且能够在上下文与主题信息之间更好地获取到语义关系和情感信息。

神经网络模型能够有效地学习到特征表示，因此，能够提高特定主题情感分类任务。Tang 等人[15]提出的 TD-LSTM 和 TC-LSTM 模型，考虑了主题信息，在特定主题的情感分类任务中，提高了分类的准确率。TC-LSTM 通过平均主题信息中包含的词向量获取主题词向量。Wang 等人[16]提出的基于注意力机制的特定主题情感分析，使注意力机制能在不同注意力被输入时，关注句子的不同部分，该模型也在特定主题情感分类任务上取得了比较好的效果。Chen 等人[17]提出的基于循环注意力记忆网络的特定主题情感分析，将循环神经网络把多个位置注意力机制进行非线性的组合，使模型能够处理更复杂的句子。Tang 等人[18]提出的基于深度记忆网络的特定主题情感分析，将内容注意力与位置注意力构建深度记忆网络，该模型能够捕获特定主题情感分类中重要的上下文单词，在处理速度和准确度上有提升。梁斌等人[19]提出的基于多注意力卷积神经网络的特定目标情感分析，融合三种注意力机制构造多注意力卷积神经网络模型，取得了比普通卷积神经网络、基于单注意力机制的卷积神经网络和基于注意力机制的 LSTM 网络更好的情感分类效果。Baziotis 等人[20]提出的基于注意力的深度 LSTM 模型，在 SemEval-2017 情感分类任务中，取得了比较好的效果。从这些研究工作中可以看出，融合特定主题和注意力的深度学习模型已经在特定主题情感分析任务中取得了很多的成果。但是这些模型没有充分融合经过特征提取后的主题信息，对于注意力的应用，也未能将多层注意力机制应用到情感分析的任务中。

融合主题信息的深度学习模型，虽然将主题信息作为单独的输入，改善了主题信息在文本中的作用，但是不能够把文本信息中影响主题的关键部分提取出来。基于注意力的深度学习模型能够获取到文本中关键信息，但是忽略了主题信息的作用。融合主题信息和单层注意力的深度学习模型，如 Wang 提出的 ATAE-LSTM[16]，将主题信息与文本信息进行连接作为输入，采用单层注意力，未能充分有效的利用主题信息和文本中的关键信息。本文构建的融合主题特征的深度注意力的 LSTM 模型，将文本信息和主题信息同时作为输入，利用双向 LSTM 共享权重的方式进行训练，将得到的文本特征和主题特征进行融合，将得到的融合特征利用深层注意力机制进行处理，有效识别文本中不同主题的情感信息。

1.2　LSTM

循环神经网络（RNN）改进了前馈神经网络，但是循环神经网络依然面临着梯度消失和梯度爆炸问题。为了解决该问题，Hochreiter 等人[21]提出和实现了长短记忆网络。在一个 LSTM 模型中，包括输入门 i^t、遗忘门 f^t、输出门 o^t 和记忆单元 c^t。输入词向量序列 $\{x_1, x_2, \cdots, x_n\}$，$x_t$ 是一个 LSTM 单元的输入，表示输入序列中一个单词的词向量。h_t 表示隐藏层向量。LSTM 中每个单元的三个门和记忆单元的计算如下：

$$X = \begin{bmatrix} h_{t-1} \\ x_t \end{bmatrix} \tag{1}$$

$$f_t = \sigma(W_f \cdot X + b_f) \tag{2}$$

$$i_t = \sigma(W_i \cdot X + b_i) \tag{3}$$

$$o_t = \sigma(W_o \cdot X + b_o) \tag{4}$$

$$c_t = f_t \odot c_{t-1} + i_t \odot tanh(W_c \cdot X + b_c) \tag{5}$$

$$h_t = o_t \odot \tanh(c_t) \tag{6}$$

其中：$W_i, W_f, W_o \in R^{d\times 2d}$ 表示权重矩阵；$b_i, b_f, b_o \in R^d$，表示 LSTM 在训练过程中学习到的偏置值；σ 表示激活函数；⊙ 表示逐个点乘积。

1.3　注意力机制

注意力机制能够使模型更多地关注文本中的重要信息。根据 LSTM 产生的隐藏层特征 $H = [h_1, \cdots, h_N]$ 构建注意力机制的输入，$H \epsilon R^{d\times N}$，其中：d 表示隐藏层的长度；N 表示输入句子的长度。注意力机制将会产生注意力权重矩阵 α 和特征表示 v。

$$u_i = tanh(W_s h_i + b_s) \tag{7}$$

$$\alpha_i = \frac{exp(u_i)}{\sum_i exp(u_i)}, \ \sum_i \alpha_i = 1 \tag{8}$$

$$v = \sum_i \alpha_i h_i \tag{9}$$

2 融合主题特征的深层注意力的 LSTM 模型

为了关注更多有价值的信息，本文提出了融合主题的深层注意力的 LSTM 模型（deeper attention LSTM with aspect embedding，AE-DATT-LSTM），利用改进后的注意力模型和两层双向 LSTM 构成，能够更好地提取主题信息和情感信息。深度注意力 LSTM 模型共包含词向量层、双向 LSTM 层、池化层、特征融合层、注意力层，其中这五层的构建过程如图 1 所示。

2.1 任务定义

对于由 n 个词组成的句子或文本 $s=\{w_1,w_2,\cdots,w_i,\cdots,w_n\}$，以下内容用文本作为句子的统称，$w_i$ 为一个文本中出现的主题词。特定主题情感分析的目标是决定文本 s 中对于主题词 w_i 的情感极性。例如，文本 “If you want a casual neighborhood bistro that has great food and excellent service, this is the place.”，对于主题“food”来说，情感极性是 Positive，但是对于主题“service”来说，情感极性是 Positive。当处理一个文本集的时，将每个词映射到低维度、连续的和实值向量，也称做词向量。所有的词向量由一个词向量矩阵 $L\in R^{d\times|V|}$，其中：d 表示词向量的维度；$|V|$ 表示词典的大小。文本中词 w_i 表示的词向量为 $w_i^s\in R^{d\times 1}$，主题词 w_j 表示的词向量为 $w_j^t\in R^{d\times 1}$，w_i 和 w_j 为词向量矩阵 L 中的一列。

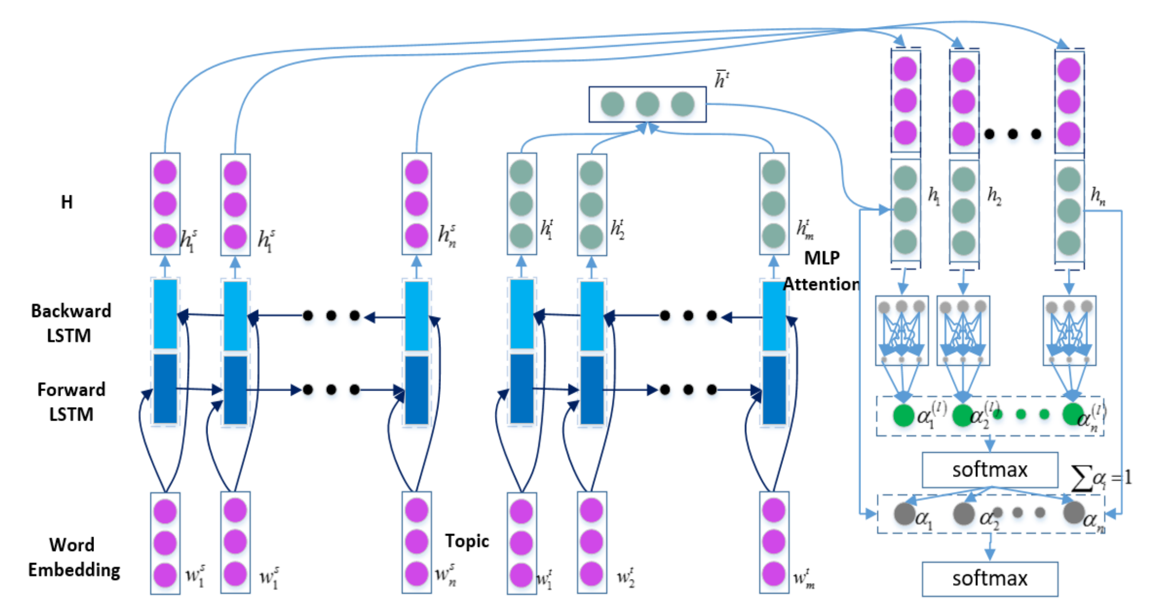


图 1 融合主题特征的深层注意力的 LSTM 模型

2.2 构建词向量输入层

融合主题特征的深层注意力的 LSTM 网络的输入由两部分构成，将文本中词的输入序列转换成词向量为 $w^s=\{w_1^s,w_2^s,\cdots,w_n^s\}$，其中，$n$ 表示文本中词的数量；将文本中的主题词转换为词向量位置 $w^t=\{w_1^t,w_2^t,\cdots,w_m^t\}$，其中，$m$ 表示文本中主题的词的数量。

2.3 构建双向 LSTM

为了能够在文本词向量与主题词向量之间更有意义的比较，本文使用共享权重的双向 LSTM 将文本和主题映射到相同向量空间。双向 LSTM 在文本层面产生的特征序列为 $H^s=\{h_1^s,h_2^s,\cdots,h_n^s\}$，在主题层面产生的特征向量为$H^t=\{h_1^t,h_2^t,\cdots,h_m^t\}$，每一部分特征向量都是由前向和后项 LSTM 连接产生。$h_i^j=\overrightarrow{h_i^j}\,||\,\overleftarrow{h_i^j}$，$h_i^j\in R^{2L}$，$j\in\{s,t\}$，其中，$||$ 表示连接操作；L 表示每个 LSTM 的大小。

2.4 池化层

本文使用一个池化层将主题特征进行聚合，进而产生单一的特征。池化层利用 *mean over time* 生成主题特征，$\overline{h^t}=\frac{1}{m}\sum_1^m h_i^t$。

2.5 特征融合层

将文本经过 LSTM 产生的特征与经过池化层将主题聚合后的特征进行连接，对每一个词形成融合特征，$h_i^{'}=h_i^s\,||\,\overline{h^t}$，$h_i^j\in R^{4L}$。

2.6 深层注意力

文本中针对特定主题的情感分析，需要更多地关注句子中的主题信息和情感词信息。因此本文引入多层 MLP 构建的深度注意力机制去提取句子中含有重要意义的主题词和情感词。

为了增强对于句子中针对主题表达情感词向量的贡献率，采用深层感知注意力机制，通过多层 MLP 实现模型。通过 l 层 MLP 深层语义注意力，产生注意力权重 α_i。

$$\alpha_i^{(1)}=RELU\left(W^{(1)}h_i'+b^{(1)}\right) \quad (10)$$

$$\vdots$$

$$\alpha_i^{(l-1)} = RELU\left(W^{(l-1)}\alpha_i^{(l-2)} + b^{(l-1)}\right) \tag{11}$$

$$\alpha_i^{(l)} = W^{(l)}\alpha_i^{(l-1)} + b^{(l)} \tag{12}$$

$$\alpha_i = softmax\left(\alpha_i^{(l)};\alpha_1^{(l)},\cdots,\alpha_n^{(l)}\right), \sum_{i=1}^{n}\alpha_i = 1 \tag{13}$$

$$h^* = \sum_{i=1}^{n}\alpha_i h_i', r \in R^{4L} \tag{14}$$

其中：$W^{(1)},W^{(2)},\cdots,W^{(l)}$，$b^{(1)},b^{(2)},\cdots,b^{(l)}$ 在训练中得到。

将 h^* 作为情感分类的特征，利用 softmax 分类器，将文本进行情感分类。

$$p = softmax\left(W_h h^* + b_h\right) \tag{15}$$

2.7　模型训练

本文的模型训练采用端到端的反向传播方式，目标函数（代价函数）采用交叉熵代价函数。将 y 作为文本中已知目标主题分配的情感类别，将 $\hat{y}$ 作为文本中目标主题的预测情感类别。模型训练的目标是最小化所有句子中已知情感类别和预测情感类别的交叉熵。

$$loss = -\sum_i\sum_j y_i^j \log \hat{y}_i^j + \lambda\|\theta\| \tag{16}$$

其中：i 表示句子的索引；j 表类别的索引，共分为三类；λ 表示 L_2 正则化，为代价函数的惩罚项；θ 表示设置的参数。

3　实验

本文将融合主题特征的深层注意力 LSTM 模型应用到特定主题情感分类任务上。在本文的实验中采用 Pennington 等人[22]提出的预训练好的 300 维 Glove 词向量进行初始化，词典的大小为 1.9 M。对于词典中未出现的词，采用均匀分布 $U(-\varepsilon,\varepsilon)$，进行随机初始化，其中 ε 设置为 0.01。对于一个文本中已知的主题，目标是确定该主题的情感极性。例如，给定一个文本“The restaurant was too expensive.”，该文本的主题词是 price，情感极性是 Negative。该情感分析模型运行在 Unbuntu 16.04 系统，运行环境为 Keras 1.2.2 和 NVIDIA GTX 1050ti 显卡。

3.1　实验数据

本文实验采用的英文数据为 SemEval-2014 Task4 数据集[11]和 SemEval-2017 Task4 数据集[12]。SemEval 是语义评测比赛任务的数据集。SemEval-2014 Task4 包含 Laptop 和 Restaurant 两个数据集。由于只有 Restaurant 数据集包含基于主题的情感极性，所以本文实验采用 Restaurant 数据集，包含顾客的评论文本，其中有 Positive、Neutral 和 Negative 三种情感，{food, price, service, ambience, anecdotes/miscellaneous}五种主题。SemEval-2017 Task4 包含 Twitter 的评论数据，该评论的主题从文本中摘取，属于主题的情感极性包含 Positive 和 Negative 两种。每条评论都包含一系列的主题和主题相关的情感极性。本文的目标是区分文本中不同主题的情感极性，包括 Positive 和 Negative 两种情感。本文实验所需要的数据统计如表 1 所示。

表 1　实验数据统计

Data Set		Positive	Neutral	Negative
Restaurant	Train	2 179	500	839
	Test	657	94	222
Twitter	Train	14 897	-	3 997
	Test	2 463	-	3 722

3.2　实验参数设置

文本中词向量的维度、主题词向量的维度和经过 LSTM 输出的特征连接后的维度都为 300。注意力权重的维度和文本中的长度是一致的。通过每个 batch 中 64 个样本的方式进行模型的训练，设置 Adam 的学习率为 0.001，设置代价函数的惩罚项为 0.001。深层注意力的层数设置为 4 层。Dropout 设置为 0.5。同时设置 LSTM 的层数为 64 层，双向 LSTM 的层数为 128 层。

3.3　实验对比

将本文提出的融合主题词向量的深层注意力 LSTM 方法 AE-DATT-LSTM 与以下六种方法在两个不同数据集上进行实验：

a）LSTM。标准的 LSTM 模型[21]不能获取到文本中任何的主题信息。虽然给予不同的主题，但得到情感极性都一致。

b）TD-LSTM。Tang 提出来的 TD-LSTM 方法[15]，使用前向和后向 LSTM 方法在主题词之前和之后提取信息。最后一个 step，模型 TD-LSTM 采用 LSTM 的输出特征来表示预测内容分类。TD-LSTM 通过将一个主题特征作为一个目标改善情感分类的表现。但是由于没有在 TD-LSTM 模型中应用注意力机制，并不能够获取文本中对于给定主题的重要词信息。

c）TC-LSTM。Tang 改进了 TD-LSTM 模型。TC-LSTM 模型[15]在原来 TD-LSTM 的基础上，将主题向量引入一个句子的特征表示。TC-LSTM 模型引入了主题连接组件，能够更好地利用主题词和文本中每个词，将其连接组成一个文本的特征表示。

d）AT-LSTM。标准的 LSTM 对于基于主题的情感分类不能够侦测到文本中的重要信息。为了解决这个问题，Wang 引入了注意力机制，提出 AT-LSTM 模型[16]，能够获取到文本中给定主题的关键部分。

e）ATAE-LSTM。AE-LSTM 中使用主题信息的方式让主题词向量在注意力权重中起到了重要的作用。为了更好地利用主题信息，Wang 改进了 AE-LSTM，提出 ATAE-LSTM[16]，将主题词向量连接到每个单词的输入向量。

f）AE-ATT-BLSTM。融合主题的单层注意力 LSTM 方法。该方法主要在特征融合层之后加一个基于单层 MLP 的注意力。

3.4　实验结果与实验分析

本文选取 SemEval-2014 Task4 中 Restaurant 数据集的训练样本作为融合主题特征的深层注意力 LSTM 方法的训练集进行

模型训练和交叉校验，采用其测试样例进行测试，得到的结果如表 2 所示。其中表 2 给出的是六种模型在 Restaurant 数据集上的三种情感极性（Positive、Negative 和 Neutral）和两种情感极性（Positive 和 Negative）的分类准确率效果。本文提出的融合主题特征的深层注意力 LSTM 模型比其他六种模型都能提高分类的准确率。引入主题特征的模型比无主题特征的模型效果要更好，引入注意力特征的比无注意力特征的效果要好。因此可以得出，主题特征和注意力机制的引入，对于基于特定主题的情感分类任务具有重要的意义。融合主题特征的深层注意力 LSTM 模型要比单引入注意力的和单引入主题的效果都要好。

表 2　六种方法在 Restaurant 数据集的分类准确率

Models	Pos/Neg/Neu	Pos/Neg
LSTM	82.0	88.3
TD-LSTM	82.6	89.1
TC-LSTM	81.9	89.2
AT-LSTM	83.1	89.6
ATAE-LSTM	84.0	89.9
AE-ATT-LSTM	85.2	90.4
AE-DATT-LSTM	86.3	90.7

本文将融合主题特征的深层注意力的 LSTM 模型在 50 次 epoch 过程中的训练过程进行分析，其中包含 accuracy、precision 和 recall 的变化趋势图。图 2 展示的是融合主题的深层 LSTM 模型在 50 个 epoch 内的训练和测试过程，包括模型中损失函数的变化，训练、交叉验证和测试中 accuracy、precision 和 recall 的变化情况。在测试阶段，当第 18 个 epoch 时，分类效果取得最优，此时的 F 值为 87.64。

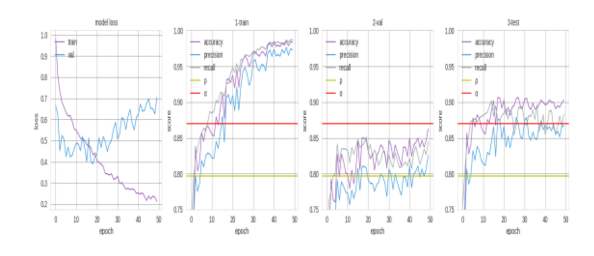


图 2　融合主题的深层注意力的 LSTM 训练过程

图 3 展示的是融合主题特征的单层注意力和深层注意力 LSTM 准确率对比。在 50 个 epoch 训练过程中，融合主题特征的深层注意力 LSTM 模型要比单层注意力 LSTM 模型的整体效果要好。图 4 展示的是单层注意力和深层注意力的 F1 值的对比。从中可以得出深层注意力 LSTM 模型要比单层注意力模型在整体上效果更佳。

为了验证本文提出的方法在不同数据集上都能取得较好的情感分类效果，将 AE-ATT-LSTM 与 AE-DATT-LSTM 模型在 SemEval-2017 Task4 中 Twitter 的评论数据集上进行实验，实验结果如表 3 所示。实验结果表明，该任务中融合主题特征的多层注意力的 LSTM 模型要比融合主题特征的单层注意力模型在分类的准确率上取得了更好的效果。

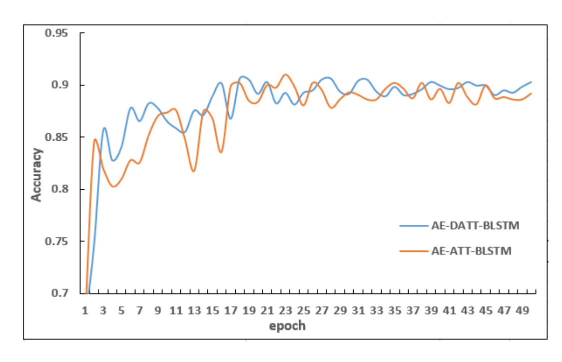


图 3　融合主题的单层与深层注意力的准确率对比

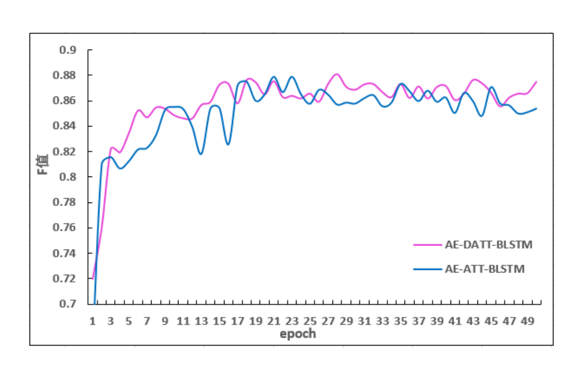


图 4　融合主题的单层与深层注意力的 F 值对比

表 3　Twitter 数据集分类准确率

Models	Positive/Negative
AE-ATT-LSTM	85.08
AE-DATT-LSTM	85.62

3.5　可视化注意力机制

为了更好地理解深层注意力模型，本文抽取出 Restaurant 数据集中的一个样本，可视化文本中上下文单词的权重。图 5 展示的是基于深层注意力 LSTM 模型的结果。文本(a)表示的主题是“food”，文本(b)表示的主题是“service”。颜色深度表示在注意力向量 α 中权重的重要等级。注意力能够动态的侦测到整个文本中的针对已知主题的情感单词，即使文本中包含有其他主题和其他情感词。

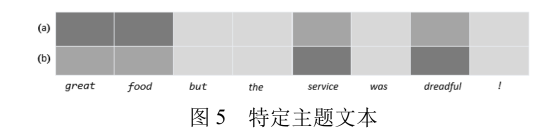


图 5　特定主题文本

4　结束语

针对特定主题情感分类任务中，大多数方法没有对主题信息和情感信息进行有效的结合，本文构建了融合主题特征的深层注意力 LSTM 方法，将主题词向量和文本词向量通过共享权重的双向 LSTM 模型进行训练，从而进行特征融合。经过深层注意力机制的处理，当给出文本中不同的主题时，该方法能够获取到文本中的不同信息。通过在不同评测集上的对比实验表

明，该方法比本文提到的其他方法在准确率和稳定性上有了进一步的提升，从而能够更好的解决特定主题情感分析任务。

本文方法未考虑词性信息，如何将词性注意力方法引入到特定主题情感分析任务中，是下一步工作的重点。本文接下来将针对词性问题进行改进。

参考文献：

[1] 王仲远，程建鹏，王浩勋，等. 短文本理解研究 [J]. 计算机研究与发展, 2016, 53 (2): 262-269.

[2] Liu Bing. Sentiment analysis and opinion mining [J]. Synthesis Lectures on Human Language Technologies, 2012, 5 (1): 1-167.

[3] Socher R, Perelygin A, Wu J, *et al.* Recursive deep models for semantic compositionality over a sentiment treebank [C]// Proc of Conference on Empirical Methods in Natural Language Processing. 2013.

[4] Tang D Y, Qin B, Liu T. Document modeling with gated recurrent neural network for sentiment classification [C]// Proc of Conference on Empirical Methods in Natural Language Processing. 2015: 1422-1432.

[5] Tai K S, Socher R, Manning S D. Improved semantic representations from tree-structured long short-term memory networks [C]// Proc of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing. 2015: 1556-1566.

[6] Minh V, Graves N H A, Kavukcuoglu K. Recurrent models of visual attention [C]// Advances in Neural Information Processing Systems. 2014: 2204-2212.

[7] Bahdanau D, Cho K, Bengio Y. Neural machine translation by jointly learning to align and translate [C]// Proc of International Conference on Learning Representations. 2015.

[8] Rush A M, Chopra S, Weston J. A neural attention model for sentence summarization [C]// Proc of Conference on Empirical Methods in Natural Language Processing. 2015: 379-389.

[9] Yang Z C, Yang D Y, Dyer C, *et al.* Hierarchical attention networks for document classification [C]// Proc of NAACL-HLT. 2016: 1480-1489.

[10] Pavlopoulos J, Malakasiotis P, Androutsopoulos I. Deeper attention to abusive user content moderation [C]// Proc of Conference on Empirical Methods in Natural Language Processing. 2017: 1136-1146.

[11] Pontiki M, Galanis D, Pavlopoulos J, *et al.* Semeval-2014 task4: aspect based sentiment analysis [C]// Proc of the 8th International Workshop on Semantic Evaluation. 2014: 27-35.

[12] Rosenthal S, Farra N, Nakov P. SemEval-2017 task4: sentiment analysis in twitter [C]// Proc of SemEval. 2017.

[13] Kiritchenko S, Zhu X D, Cherry C, *et al.* NRC-Canada-2014: detecting aspects and sentiment in customer reviews [C]// Proc of the 8th International Workshop on Semantic Evaluation. 2014: 437-442.

[14] Nguyen T H, Shirai K. PhraseRNN: phrase recursive neural network for aspect-based sentiment analyisi [C]// Proc of Conference on Empirical Methods in Natural Language Processing. 2015: 2509-2514.

[15] Tang D Y, Qin B, Feng X C, *et al.* Effective LSTMs for target-dependent sentiment classification [C]// Proc of the 26th International Conference on Computational Linguistics. 2016: 3298-3307.

[16] Wang Y Q, Huang M, Zhao L, *et al.* Attention-based LSTM for aspect-level sentiment classification [C]// Proc of Conference on Empirical Methods in Natural Language Processing. 2016: 606-615.

[17] Chen P, Sun Z Q, Bing L D, *et al.* Recurrent attention network on memory for aspect sentiment analysis [C]// Proc of Conference on Empirical Methods in Natural Language Processing. 2017: 463-472.

[18] Tang D Y, Qin B, Liu T. Aspect level sentiment classification with deep memory network [C]// Proc of Conference on Empirical Methods in Natural Language Processing. 2016: 214-224.

[19] 梁斌，刘全，徐进，等. 基于多注意力卷积神经网络的特定目标情感分析 [J]. 计算机研究与发展, 2017, 54 (8): 1724-1735.

[20] Baziotis C, Pelekis N, Doulkeridis C. DataStories at SemEval-2017 task 4: deep LSTM with attention for message-level and topic-based sentiment analysis [C]// Proc of the 11th International Workshop on Semantic Evaluations. 2017: 747-754.

[21] Hochreiter S, Schmidhuber J. Long short-term memory [J]. Neural Computation, 1997, 9 (8): 1735-1780.

[22] Pennington J, Socher R, Manning C D. GloVe: global vectors for word representation [C]// Proc of Conference on Empirical Methods in Natural Language Processing. 2014: 1532-1543.